



# An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors

José Henrique de Morais Goulart, Gérard Favier

## ► To cite this version:

José Henrique de Morais Goulart, Gérard Favier. An algebraic solution for the Candecomp/PARAFAC decomposition with circulant factors. SIAM Journal on Matrix Analysis and Applications, 2014, 35 (4), pp.1543-1562. hal-00967263v2

**HAL Id: hal-00967263**

**<https://hal.science/hal-00967263v2>**

Submitted on 17 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AN ALGEBRAIC SOLUTION FOR THE CANDECOMP/PARAFAC DECOMPOSITION WITH CIRCULANT FACTORS

J. H. DE M. GOULART<sup>\*†</sup> AND G. FAVIER<sup>\*</sup>

**Abstract.** The Candecomp/PARAFAC decomposition (CPD) is an important mathematical tool used in several fields of application. Yet, its computation is usually performed with iterative methods which are subject to reaching local minima and to exhibiting slow convergence. In some practical contexts, the data tensors of interest admit decompositions constituted by matrix factors with particular structure. Often, such structure can be exploited for devising specialized algorithms with superior properties in comparison with general iterative methods. In this paper, we propose a novel approach for computing a circulant-constrained CPD (CCPD), i.e., a CPD of a hypercubic tensor whose factors are all circulant (and possibly tall). To this end, we exploit the algebraic structure of such tensor, showing that the elements of its frequency-domain counterpart satisfy homogeneous monomial equations in the eigenvalues of square circulant matrices associated with its factors, which we can therefore estimate by solving these equations. Then, we characterize the sets of solutions admitted by such equations under Kruskal's uniqueness condition. Simulation results are presented, validating our approach and showing that it can help avoiding typical disadvantages of iterative methods.

**Key words.** Candecomp/PARAFAC decomposition, canonical polyadic decomposition, tensor decomposition, circulant matrices, homogeneous monomial equations

**AMS subject classifications.** 15A69

**1. Introduction.** The Candecomp/PARAFAC or canonical polyadic decomposition (CPD) [19, 6] is a mathematical tool of great utility in many practical applications [4, 22]. Basically, it consists of a generalization of bilinear decompositions of matrices to higher-order tensors. In addition, unlike its bilinear counterparts, it is essentially unique under mild conditions. Therefore, it is naturally well suited for the analysis of data sets constituted by observations of a function of multiple discrete indices, as encountered in signal processing [30, 16, 15, 17], data mining [2, 27] and biomedical engineering [1]; see [4, 22] for other examples.

The attractive properties of the CPD are, however, accompanied by difficulties. Namely, its computation is generally a difficult non-convex optimization problem and, furthermore, determining the minimal number of terms that are necessary to decompose a tensor, which corresponds to its rank, is NP-hard [20]. Notwithstanding these difficulties, there are several methods which try to fit a best rank- $R$  approximate CPD model to the data tensor of interest [34, 22].

In the general case, when no assumption is made about the matrix factors involved in the decomposition, one usually resorts to iterative methods for estimating them. Among these methods, the so-called alternating least-squares (ALS), which was originally and independently proposed in [6, 19], is the most used. Another popular algorithm is the classical Levenberg-Marquardt optimization method for nonlinear least-squares problems [26]. However, given the nature of the CPD computation problem, which is generally formulated as the minimization of a least-squares objective function, iterative methods are evidently subject to convergence towards local minima. Moreover, they often suffer from slow convergence speed, and therefore may require a large amount of computing time.

---

<sup>\*</sup>Laboratoire I3S, Université de Nice Sophia-Antipolis, CNRS, 2000 route des Lucioles, Les Algorithmes/Euclide B, 06900 Sophia-Antipolis, France (goulart@i3s.unice.fr, favier@i3s.unice.fr).

<sup>†</sup>Sponsored by CNPq-Brazil (individual grant 245358/2012-9).

Conceivably, when the matrix factors of a CPD are known to have special structure, as, e.g., Toeplitz, circulant or Vandermonde, one could take this into account to formulate specialized algorithms which avoid the disadvantages of iterative methods. Indeed, such strategy has been followed in some works, motivated by the pertinence of structural constraints in several practical contexts. For instance, [21] proposes a non-iterative method for the estimation of a third-order CPD in which a banded circulant factor is involved. Similarly, [31] develops methods which can recover in a non-iterative fashion banded (and possibly Toeplitz or Hankel) factors of a tensor. Another recent example of a structurally constrained CPD estimation method is [32], where Vandermonde matrix factors are considered. As examples of applications where some structurally constrained CPD arises, we can mention the blind identification of single-input single-output (SISO) finite impulse response (FIR) channels via the use of high-order statistics [17] and the estimation of parameters of Wiener-Hammerstein nonlinear systems from the kernels of equivalent Volterra systems [21]. Constrained structures have also been exploited for estimating other tensor models. For instance, in [14] the third-order core tensors of block Tucker models are characterized by matrix slices having Hankel and Vandermonde forms, which is taken into account for deriving a specialized parameter estimation algorithm.

In this paper, we develop a novel algebraic approach for computing a circulant-constrained CPD (CCPD), which consists of a CPD of a hypercubic tensor whose factors are all circulant (and possibly tall). This approach relies on the resolution of a system of homogeneous monomial equations, which are directly obtained from the multidimensional Fourier transform of the tensor to be decomposed and follow from the well known Fourier eigenstructure property of circulant matrices. In the (ideal) noiseless setting, this yields an exact solution for the decomposition. We also show how to specialize this approach to a symmetric CCPD, whose factors are all identical. In contrast to the existing methods [21, 31], which can be used to estimate circulant factors of a CPD, our approach does not impose additional constraints over these factors (apart from Kruskal's standard uniqueness condition) and is capable of taking into account the entire structure of a CCPD, symmetric or not. These differences come from the fact that [21, 31] rely on subspace analysis of a matrix unfolding of the tensor of interest. We note also that, due to the permutation ambiguity which is inherent to the CPD, our approach is as well valid for estimating factors with any structure that can be transformed into a Toeplitz-circulant structure via post-multiplication by a permutation matrix, as, e.g., "Hankel-circulant" matrices. As another contribution, we formulate a version of the ALS algorithm which constrains the estimated factors of a CPD to be circulant, and thus is more appropriate as a reference for comparison with our algebraic approach.

The contents of this paper are organized as follows. Section 2 briefly reviews important concepts and definitions related to tensors, the CPD, its basic properties and the main existing methods for its computation. Then, the proposed approach is developed in §3, after recalling the fundamental property of circulant matrices on which it is based. Some simple illustrative examples of the application of our approach are provided in §4, for enhancing the understanding of the developed analysis. Next, §5 presents and discusses some simulation results through which the proposed approach is compared with several other methods. Finally, concluding remarks are given in §6.

**Notational conventions:** Matrices and column vectors are denoted by boldface capital and boldface lowercase letters, respectively. The columns of a  $M \times R$  matrix  $\mathbf{A}$  are denoted by  $\mathbf{a}_1, \dots, \mathbf{a}_R$ , and it is said to be tall if  $M > R$ . We denote the transpose,

Hermitian transpose and Moore-Penrose pseudoinverse of  $\mathbf{A}$  by, respectively,  $\mathbf{A}^T$ ,  $\mathbf{A}^H$  and  $\mathbf{A}^\dagger$ , and the  $R \times R$  identity matrix by  $\mathbf{I}_R$ . The operator  $\text{vec}(\cdot)$  maps  $\mathbf{A} \in \mathbb{C}^{M \times R}$  into  $\mathbf{a} = [\mathbf{a}_1^T \ \dots \ \mathbf{a}_R^T]^T \in \mathbb{C}^{RM}$ , and  $\text{diag}(\cdot)$  maps a vector  $\mathbf{v} \in \mathbb{C}^M$  into an  $M \times M$  matrix which contains its elements on the main diagonal and zeros everywhere else. Given  $\mathbf{v}, \mathbf{w} \in \mathbb{C}^M$ , the Hadamard product  $\mathbf{v} \odot \mathbf{w} \in \mathbb{C}^M$  yields their element-wise multiplication. The Kronecker product between matrices  $\mathbf{A} \in \mathbb{C}^{M_1 \times R_1}$  and  $\mathbf{B} \in \mathbb{C}^{M_2 \times R_2}$  is denoted by  $\mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{M_1 M_2 \times R_1 R_2}$ , while their Khatri-Rao (i.e., columnwise Kronecker) product, which is only defined if  $R_1 = R_2 = R$ , is denoted by  $\mathbf{A} \diamond \mathbf{B} \in \mathbb{C}^{M_1 M_2 \times R}$ . We use  $\delta_{i_1, \dots, i_P}$  to denote the generalized Kronecker delta, i.e.,  $\delta_{i_1, \dots, i_P} = 1$  if  $i_1 = i_2 = \dots = i_P$  and  $\delta_{i_1, \dots, i_P} = 0$  otherwise.

## 2. Review of the CP decomposition.

**2.1. Basic definitions.** Formally, a tensor of order  $P$  is an element of a tensor product among  $P$  vector spaces [25]. However, once a certain basis is chosen for each of these spaces, a tensor can be represented by a multidimensional array of numbers from a field (usually  $\mathbb{R}$  or  $\mathbb{C}$ ). We denote a  $P$ th order tensor of dimensions  $M_1 \times \dots \times M_P$  over  $\mathbb{C}$  by a capital calligraphic letter  $\mathcal{T}$ , and its corresponding space by  $\mathbb{C}^{M_1 \times \dots \times M_P}$ . Its elements are denoted by lowercase letters,  $t_{m_1, \dots, m_P}$ , having indices  $m_1, \dots, m_P$  which satisfy  $m_j \in \{1, \dots, M_j\}$ , or, alternatively, by the notation  $[\mathcal{T}]_{m_1, \dots, m_P}$ .

Each geometric dimension of a tensor is called a mode. Given a tensor  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ , for each mode  $p$  we can arrange its elements in a matrix unfolding  $\mathbf{T}_p \in \mathbb{C}^{M_p \times M_P \dots M_{p+1} M_{p-1} \dots M_1}$  with elements  $[\mathbf{T}_p]_{m_p, j} = t_{m_1, \dots, m_P}$ , where [22]

$$j = 1 + \sum_{\substack{q=1 \\ q \neq p}}^P (m_q - 1) \prod_{\substack{r=1 \\ r \neq p}}^{q-1} M_r.$$

A  $P$ th order tensor  $\mathcal{T}$  whose all modes have the same dimension  $M$  is called hypercubic and is denoted by  $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$ , where  $M^{(P)}$  is a shorthand for  $M \times \dots \times M$ , with  $P$  occurrences of  $M$ . A hypercubic tensor  $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$  is said to be symmetric whenever its elements are invariant to any permutation  $\pi(m) = (\pi_1, \dots, \pi_P)$  of their indices  $m = (m_1, \dots, m_P)$ , i.e.,  $t_{m_1, \dots, m_P} = t_{\pi_1, \dots, \pi_P}$ . The  $n$ -mode product [7] of  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$  by a matrix  $\mathbf{A} \in \mathbb{C}^{K \times M_n}$ , denoted by  $\mathcal{T} \times_n \mathbf{A}$ , yields a  $M_1 \times \dots \times M_{n-1} \times K \times M_{n+1} \times \dots \times M_P$  tensor  $\mathcal{S}$  such that

$$(2.1) \quad s_{m_1, \dots, m_{n-1}, k, m_{n+1}, \dots, m_P} = \sum_{m_n=1}^{M_n} a_{k, m_n} t_{m_1, \dots, m_n, \dots, m_P} \quad \Leftrightarrow \quad \mathbf{S}_n = \mathbf{A} \mathbf{T}_n.$$

From (2.1), it can be shown that, for  $m \neq n$  and  $\mathbf{A}, \mathbf{B}$  of appropriate dimensions,

$$(2.2) \quad (\mathcal{T} \times_n \mathbf{A}) \times_m \mathbf{B} = \mathcal{T} \times_n (\mathbf{B} \mathbf{A}) \quad \text{and} \quad (\mathcal{T} \times_m \mathbf{A}) \times_n \mathbf{B} = (\mathcal{T} \times_n \mathbf{B}) \times_m \mathbf{A}$$

hold. Another important operator is the outer product (or tensor product [25]), which is defined as follows. Let  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$  and  $\mathcal{S} \in \mathbb{C}^{K_1 \times \dots \times K_Q}$  be tensors of orders  $P$  and  $Q$ , respectively. Their outer product, denoted by  $\mathcal{T} \circ \mathcal{S}$ , yields a tensor  $\mathcal{V} \in \mathbb{C}^{M_1 \times \dots \times M_P \times K_1 \times \dots \times K_Q}$  (of order  $P + Q$ ) that can be expressed in scalar form as

$$v_{m_1, \dots, m_P, k_1, \dots, k_Q} = t_{m_1, \dots, m_P} s_{k_1, \dots, k_Q}.$$

Note that this definition includes as a particular case the well known outer product between two vectors, which yields a matrix.

Let  $\omega_M \triangleq \exp(j2\pi/M)$  and  $\mathbf{F}_M$  be the  $M \times M$  Fourier matrix, given by

$$(2.3) \quad \mathbf{F}_M = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_M & \omega_M^2 & \dots & \omega_M^{M-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega_M^{M-2} & \omega_M^{M-4} & \dots & \omega_M^2 \\ 1 & \omega_M^{M-1} & \omega_M^{M-2} & \dots & \omega_M \end{bmatrix}.$$

The multidimensional discrete Fourier transform (MDFT) of  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ , which we denote by  $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$ , is defined as

$$(2.4) \quad \mathcal{Y} = \mathcal{T} \times_1 \mathbf{F}_{M_1}^H \times_2 \mathbf{F}_{M_2}^H \dots \times_P \mathbf{F}_{M_P}^H \in \mathbb{C}^{M_1 \times \dots \times M_P}.$$

In scalar form, we have

$$y_{k_1, \dots, k_P} = \left( \prod_{p=1}^P \frac{1}{\sqrt{M_p}} \right) \sum_{m_1=1}^{M_1} \dots \sum_{m_P=1}^{M_P} t_{m_1, \dots, m_P} \omega_{M_1}^{-(m_1-1)(k_1-1)} \dots \omega_{M_P}^{-(m_P-1)(k_P-1)},$$

for  $k_p \in \{1, \dots, M_p\}$ ,  $p \in \{1, \dots, P\}$ . Analogously, we define the inverse multidimensional discrete Fourier transform (IMDFT) of  $\mathcal{Y} \in \mathbb{C}^{M_1 \times \dots \times M_P}$  as

$$(2.5) \quad \mathcal{T} = \mathcal{Y} \times_1 \mathbf{F}_{M_1} \times_2 \mathbf{F}_{M_2} \dots \times_P \mathbf{F}_{M_P} \in \mathbb{C}^{M_1 \times \dots \times M_P},$$

and denote it by  $\mathcal{T} = \text{IMDFT}\{\mathcal{Y}\}$ . In scalar form, this relation can be written as

$$t_{m_1, \dots, m_P} = \left( \prod_{p=1}^P \frac{1}{\sqrt{M_p}} \right) \sum_{k_1=1}^{M_1} \dots \sum_{k_P=1}^{M_P} y_{k_1, \dots, k_P} \omega_{M_1}^{(m_1-1)(k_1-1)} \dots \omega_{M_P}^{(m_P-1)(k_P-1)},$$

with  $m_p \in \{1, \dots, M_p\}$ ,  $p \in \{1, \dots, P\}$ . Note that, due to (2.2), the computation of (2.4) or (2.5) can be done with  $P$  matrix multiplications, whose ordering is irrelevant.

**2.2. The CP decomposition.** The CP decomposition (CPD) or CP model of a tensor  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$  is defined by

$$(2.6) \quad \mathcal{T} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \dots \circ \mathbf{a}_r^{(P)},$$

where  $\mathbf{a}_r^{(p)}$  is the  $r$ -th column of the matrix factor  $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$ . The minimal value of  $R$  such that  $\mathcal{T}$  can be written as in (2.6) is called the rank of  $\mathcal{T}$ . It is clear from (2.6) that the CPD can be seen as a generalization of the decomposition of a matrix into a sum of rank-one matrices; in fact, a  $P$ th order tensor is said to have rank one precisely when it can be written as an outer product of  $P$  vectors. Also, as in the matrix case, every tensor admits a decomposition of the form (2.6) [9]. Using the definition of the outer product, the CPD can be expressed in scalar form as

$$(2.7) \quad t_{m_1, \dots, m_P} = \sum_{r=1}^R \prod_{p=1}^P a_{m_p, r}^{(p)},$$

where  $a_{m, r}^{(p)}$  is the  $m$ -th element of  $\mathbf{a}_r^{(p)}$ . By defining the  $P$ th order identity tensor  $\mathcal{I} \in \mathbb{C}^{R^{(P)}}$ , whose elements satisfy  $i_{r_1, \dots, r_P} = \delta_{r_1, \dots, r_P}$ , from (2.7) we can deduce that

$$(2.8) \quad \mathcal{T} = \mathcal{I} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_P \mathbf{A}^{(P)}.$$

Here, we will also use the convenient Kruskal's notation [23], expressing (2.8) as

$$(2.9) \quad \mathcal{T} = \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)} \rrbracket.$$

A useful fact is that the unfolding  $\mathbf{T}_p$  of  $\mathcal{T}$  can be written as

$$(2.10) \quad \mathbf{T}_p = \mathbf{A}^{(p)} \left( \mathbf{A}^{(P)} \diamond \dots \diamond \mathbf{A}^{(p+1)} \diamond \mathbf{A}^{(p-1)} \diamond \dots \diamond \mathbf{A}^{(1)} \right)^T.$$

Finally, if a CPD of a symmetric tensor  $\mathcal{S}$  is constituted by identical factors, i.e.,

$$(2.11) \quad \mathcal{S} = \llbracket \mathbf{A}, \dots, \mathbf{A} \rrbracket \in \mathbb{C}^{M^{(P)}},$$

where  $\mathbf{A} \in \mathbb{C}^{M \times R_s}$ , it is called a symmetric CPD of  $\mathcal{S}$  [8, 3]. If  $R_s$  is the minimal value for which (2.11) holds, it is called the symmetric rank of  $\mathcal{S}$ .

**2.3. Uniqueness of the CPD.** The CPD (2.9) is said to be essentially unique if any other CPD of  $\mathcal{T}$  with factors  $\mathbf{B}^{(p)} \in \mathbb{C}^{M_p \times R}$ , i.e.,  $\mathcal{T} = \llbracket \mathbf{B}^{(1)}, \dots, \mathbf{B}^{(P)} \rrbracket$ , is such that  $\mathbf{B}^{(p)} = \mathbf{A}^{(p)} \mathbf{\Pi} \mathbf{\Delta}_p$ , where  $\mathbf{\Pi} \in \mathbb{R}^{R \times R}$  is a permutation matrix and the matrices  $\mathbf{\Delta}_p \in \mathbb{R}^{R \times R}$ ,  $p \in \{1, \dots, P\}$ , are diagonal and satisfy  $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$ . Basically, this definition allows us to disregard the trivial column permutation and scaling ambiguities that are inherent to the model. With it at hand, we can state a fundamental result on the uniqueness of the CPD, which was first established by Kruskal [23] for third-order tensors and then generalized by Sidiropoulos and Bro [29] for any order. Before, however, another concept must be introduced: the  $k$ -rank (named after Kruskal) of a matrix  $\mathbf{A}$ , denoted by  $k_{\mathbf{A}}$ , is the largest integer such that every set containing  $k_{\mathbf{A}}$  columns of  $\mathbf{A}$  is linearly independent [23].

**THEOREM 2.1 ([29]).** *Let  $\mathcal{T} \in \mathbb{C}^{M_1 \times \dots \times M_P}$  be given by (2.9), where  $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$ , with  $P > 2$  and  $R > 1$ . If the matrix factors  $\mathbf{A}^{(p)}$  are such that their  $k$ -ranks satisfy  $k_{\mathbf{A}^{(1)}} + \dots + k_{\mathbf{A}^{(P)}} \geq 2R + P - 1$ , then the CPD is essentially unique.*

It should be noted that the condition involved in Theorem 2.1 is sufficient but not necessary. Note also that, unlike matrix decompositions, whose uniqueness can only be established by imposing special constraints—such as, e.g., orthogonality—on their factors, the CPD requires much less constraining assumptions.

**2.4. CPD computation.** Given a data tensor  $\mathcal{D} \in \mathbb{C}^{M_1 \times \dots \times M_P}$ , the computation of a CPD of  $\mathcal{D}$  is generally stated as searching for factors  $\mathbf{A}^{(p)} \in \mathbb{C}^{M_p \times R}$ ,  $p \in \{1, \dots, P\}$ , that jointly minimize an error criterion, which is often formulated as

$$(2.12) \quad J(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)}) = \left\| \mathcal{D} - \mathcal{T}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)}) \right\|_F^2,$$

where  $\mathcal{T}(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)})$  is given by (2.9) and  $\|\cdot\|_F$  stands for the Frobenius norm, defined as  $\|\mathcal{T}\|_F \triangleq \sqrt{\text{vec}(\mathbf{T}_p)^H \text{vec}(\mathbf{T}_p)}$ , for any  $p \in \{1, \dots, P\}$ . The choice of  $R$  is clearly critical, but in practice it is usually unknown. Furthermore, computing  $R$  is NP-hard [20]. Consequently, a CPD is often computed by searching for the best rank- $R$  approximation of  $\mathcal{T}$ , given a choice of  $R$ . Yet, this problem is ill-posed in general, because a tensor may not have a best rank- $R$  approximation [12]. In that case, one can find factors which yield an arbitrarily low approximation error, although their columns contain meaningless elements with very large magnitude [28, 33].

The cost function (2.12) is clearly nonlinear in the parameters, and thus the solution is, in general, computed iteratively. To perform this task, the most commonly

used algorithm is the alternating least-squares (ALS) [19]. Basically, it consists in sequentially updating the estimate of each factor, fixing the other ones to their previous estimates. Hence, the original nonlinear problem is tackled by considering a sequence of linear ones, which can be solved using the least-squares estimator. Although this strategy is conceptually simple and has led to useful results in many applications, there is no proof of its convergence. Nonlinear optimization schemes, such as the Levenberg-Marquardt (LM) method, are also often employed [9, 34]. Unlike the ALS, techniques of this nature rely on a solid mathematical framework, but may also lead to local minima and/or converge very slowly. It should be emphasized that both these approaches rely on an *a priori* choice of  $R$ , which can be set in practice by applying techniques such as those of [5, 11]. Moreover, they are subject to computing spurious factors if the data tensor does not admit a best rank- $R$  approximation.

**3. Algebraic solution for a circulant-constrained CPD.** In the following, we provide a formulation of the problem addressed in this work, namely, the estimation of a CCPD. Then, we describe the proposed approach, which is based on the resolution of a system of monomial equations that can be obtained through a multilinear transformation of the tensor to be decomposed, provided that its rank is known *a priori*. After that, we study the solutions admitted by such equations. Finally, we specialize our analysis to the case where the CCPD is symmetric.

First, however, we briefly review some important properties of circulant matrices.

**DEFINITION 3.1.** A matrix  $\mathbf{C} \in \mathbb{C}^{M \times R}$  is said to be *circulant with generating vector*  $\mathbf{c} = [c_0 \ c_1 \ \dots \ c_{M-1}]^T \in \mathbb{C}^M$ , which is denoted by  $\mathbf{C} = \text{circ}_R(\mathbf{c})$ , if

$$(3.1) \quad \mathbf{C} = [\mathbf{c} \ \mathbf{\Pi}_M \mathbf{c} \ \dots \ \mathbf{\Pi}_M^{R-1} \mathbf{c}],$$

where  $\mathbf{\Pi}_M$  is the following  $M \times M$  permutation matrix:

$$(3.2) \quad \mathbf{\Pi}_M = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}.$$

Although no restriction has been placed over the dimensions  $M$  and  $R$ , in the sequel we consider only square and tall matrices, as will be explained in §3.1. Square circulant matrices possess a very special property that is well known: they all share the same eigenvectors, which are simply the columns of the Fourier matrix [18]. Since these columns form an orthogonal basis for  $\mathbb{C}^M$ , it follows that any circulant matrix  $\mathbf{C} \in \mathbb{C}^{M \times M}$  can be diagonalized by the Fourier matrix, or, equivalently,

$$(3.3) \quad \mathbf{C} = \mathbf{F}_M \mathbf{\Lambda} \mathbf{F}_M^H,$$

where  $\mathbf{F}_M$  is defined in (2.3) and  $\mathbf{\Lambda}$  is a  $M \times M$  diagonal matrix containing the eigenvalues of  $\mathbf{C}$ . Another direct consequence of this particular eigenstructure is that, if we rewrite (3.3) as  $\mathbf{F}_M^H \mathbf{C} = \mathbf{\Lambda} \mathbf{F}_M^H$ , then from the first column of the latter identity we have that  $\mathbf{F}_M^H \mathbf{c} = \mathbf{\lambda} / \sqrt{M}$ , where  $\mathbf{\lambda}$  is such that  $\text{diag}(\mathbf{\lambda}) = \mathbf{\Lambda}$ . In other words,  $\mathbf{c}$  and  $\mathbf{\lambda} / \sqrt{M}$  form a DFT pair. As property (3.3) applies only for square matrices, we now introduce a definition which will facilitate the treatment of tall circulant factors.

**DEFINITION 3.2.** Let  $\mathbf{C} = \text{circ}_R(\mathbf{c}) \in \mathbb{C}^{M \times R}$  with  $M \geq R$ . We define the *circulant completion* of  $\mathbf{C}$  as the square matrix  $\check{\mathbf{C}} \in \mathbb{C}^{M \times M}$  given by

$$\check{\mathbf{C}} = \begin{cases} [\mathbf{C} \ \mathbf{\Pi}_M^R \mathbf{c} \ \mathbf{\Pi}_M^{R+1} \mathbf{c} \ \dots \ \mathbf{\Pi}_M^{M-1} \mathbf{c}], & \text{if } M > R \\ \mathbf{C}, & \text{if } M = R, \end{cases}$$



where  $\mathbf{\Pi}_M$  is as defined by (3.2).

From the above definition, it follows that any  $M \times R$  circulant matrix with  $M \geq R$  is such that its circulant completion can be decomposed as in (3.3).

By noting that  $\mathbf{\Pi}_M$  itself is circulant, we can also state the following useful property, whose proof will be omitted due to the lack of space.

**PROPOSITION 3.3.** *Let  $\mathbf{\Pi}_M$  be as defined by (3.2). Then,  $\mathbf{\Pi}_M^k = \mathbf{\Pi}_M^{(k)_M}$  for all  $k \in \mathbb{Z}$ , where  $(\cdot)_M$  denotes the modulo  $M$  operator.<sup>1</sup> Consequently, there exist only  $M$  distinct matrices of the form  $\mathbf{\Pi}_M^k$ , namely,  $\mathbf{\Pi}_M^0, \mathbf{\Pi}_M, \dots, \mathbf{\Pi}_M^{M-1}$ . Moreover,  $\mathbf{\Pi}_M^k$  admits the eigendecomposition*

$$(3.4) \quad \mathbf{\Pi}_M^k = \mathbf{F}_M \text{diag} \left( 1, \omega_M^{-k}, \dots, \omega_M^{-k(M-1)} \right) \mathbf{F}_M^H.$$

**3.1. General CCPD.** Let  $\mathcal{T}$  be a hypercubic  $P$ th order tensor of dimensions  $M \times \dots \times M$ . Suppose that  $\mathcal{T}$  admits a CCPD, i.e., that there exist  $P$  complex circulant matrices  $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$ ,  $p \in \{1, \dots, P\}$ , such that  $\mathcal{T}$  can be written as

$$(3.5) \quad \mathcal{T} = \llbracket \mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

Our goal is then to determine the circulant factors of (3.5), up to trivial ambiguities.

A distinguishing feature of the CCPD (3.5) is that its rank is upper bounded by  $M$ , as opposed to the general CPD (2.9), whose rank can exceed the dimensions of the tensor [22]. Such conclusion can be reached by noting that, if the factors  $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$  in (3.5) are such that  $R > M$ , then from Definition 3.1 and Proposition 3.3 we have  $\mathbf{c}_{M+m}^{(p)} = \mathbf{c}_m^{(p)}$  for all  $p$  and  $m \geq 1$ , where  $\mathbf{c}_m^{(p)}$  denotes the  $m$ th column of  $\mathbf{C}^{(p)}$ . Clearly, this “joint repetition” would imply the existence of linearly dependent terms in the decomposition, and therefore  $R$  would not be minimal. Because of this fact, in what follows we assume that  $M \geq R$ , without loss of generality.

**3.1.1. Derivation of the monomial equations.** The first step in the derivation of our approach consists in showing how property (3.3) can lead to a set of monomial equations from a hypercubic tensor  $\mathcal{T}$  which is known to admit a CCPD. Taking the MDFT of  $\mathcal{T}$ , we have

$$(3.6) \quad \mathcal{Y} = \mathcal{T} \times_1 \mathbf{F}_M^H \times_2 \mathbf{F}_M^H \cdots \times_P \mathbf{F}_M^H \in \mathbb{C}^{M^{(P)}}.$$

Applying properties (2.2) and (2.8), it can be deduced from (3.5) and (3.6) that

$$(3.7) \quad \mathcal{Y} = \llbracket \mathbf{F}_M^H \mathbf{C}^{(1)}, \dots, \mathbf{F}_M^H \mathbf{C}^{(P)} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

But, since each  $\mathbf{C}^{(p)} \in \mathbb{C}^{M \times R}$  is a circulant matrix with  $M \geq R$ , we can decompose its circulant completion as  $\check{\mathbf{C}}^{(p)} = \mathbf{F}_M \mathbf{\Lambda}_p \mathbf{F}_M^H$ , where  $\mathbf{\Lambda}_p = \text{diag}(\boldsymbol{\lambda}_p)$  is the diagonal matrix of eigenvalues of  $\check{\mathbf{C}}^{(p)}$ . Therefore,

$$(3.8) \quad \begin{aligned} \mathbf{F}_M^H \check{\mathbf{C}}^{(p)} &= \text{diag}(\boldsymbol{\lambda}_p) \mathbf{F}_M^H \\ &= [\boldsymbol{\lambda}_p \odot \mathbf{f}_{M,1}^* \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,2}^* \quad \dots \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,M}^*], \end{aligned}$$

where  $\mathbf{f}_{M,m}$  is the  $m$ -th column of  $\mathbf{F}_M = \mathbf{F}_M^T$  and the  $*$  superscript stands for complex conjugation. Thus, considering the first  $R$  columns of (3.8), we have

$$(3.9) \quad \mathbf{F}_M^H \mathbf{C}^{(p)} = [\boldsymbol{\lambda}_p \odot \mathbf{f}_{M,1}^* \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,2}^* \quad \dots \quad \boldsymbol{\lambda}_p \odot \mathbf{f}_{M,R}^*].$$

---

<sup>1</sup> $\forall k \in \mathbb{Z}, M \in \mathbb{N}, (k)_M = m \Leftrightarrow k = lM + m$  for some  $l \in \mathbb{Z}$ , with  $0 \leq m < M$ .



In particular, as already mentioned, the first column of the identity (3.9) means that

$$(3.10) \quad \text{DFT} \left\{ \mathbf{c}^{(p)} \right\} = \frac{\lambda_p}{\sqrt{M}} \iff \text{IDFT} \left\{ \frac{\lambda_p}{\sqrt{M}} \right\} = \mathbf{c}^{(p)},$$

where  $\mathbf{c}^{(p)}$  is the generating vector of  $\mathbf{C}^{(p)}$ .

Relations (3.7) and (3.9) provide the key for deriving the equations that we shall use to estimate the CCPD (3.5). More specifically, by substituting (3.9) into (3.7) and resorting to the scalar form (2.7) of the CPD, the elements of  $\mathcal{Y}$  are seen to satisfy

$$(3.11) \quad y_{m_1, \dots, m_P} = \left( \frac{1}{\sqrt{M}} \right)^P \prod_{p=1}^P \lambda_{p, m_p} \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)},$$

where  $\lambda_{p, m}$  is the  $m$ -th element of  $\lambda_p$ . Hence, the elements of the frequency-domain counterpart of  $\mathcal{T}$  satisfy  $P$ th order monomial equations whose unknowns are the eigenvalues of the circulant completions of its matrix factors.

**REMARK 3.4.** *It is interesting to note that the above development can be generalized to handle any CPD  $\mathcal{T} = [\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(P)}]$  such that each  $\mathbf{A}^{(p)} \in \mathbb{C}^{M \times R}$  can be completed to form a square matrix  $\tilde{\mathbf{A}}^{(p)} \in \mathbb{C}^{M \times M}$  satisfying  $\tilde{\mathbf{A}}^{(p)} = \mathbf{V}^{(p)} \mathbf{\Lambda}_p \mathbf{Z}^{(p)}$ , where  $\mathbf{V}^{(p)}, \mathbf{Z}^{(p)} \in \mathbb{C}^{M \times M}$  are known a priori,  $\mathbf{V}^{(p)}$  is nonsingular and  $\mathbf{\Lambda}_p$  is diagonal (and unknown). In that case, one computes  $\mathcal{Y} = \mathcal{T} \times_1 (\mathbf{V}^{(1)})^{-1} \dots \times_P (\mathbf{V}^{(P)})^{-1}$ , whose elements are then given by  $y_{m_1, \dots, m_P} = \prod_{p=1}^P \lambda_{p, m_p} \sum_{r=1}^R [\mathbf{Z}^{(1)}]_{m_1, r} \dots [\mathbf{Z}^{(P)}]_{m_P, r}$ .*

### 3.1.2. Characterization of the derived equations and of its solutions.

The set of equations (3.11) for  $m_p \in \{1, \dots, M\}$ ,  $p \in \{1, \dots, P\}$ , might lead to a solution for the computation of the CCPD (3.5), if we can compute the eigenvalues of the circulant completion associated with each factor. To this end, it is important to first determine which equations of the form (3.11) are relevant, since the summation of complex exponentials can vanish for some multi-indices  $(m_1, \dots, m_P)$ , which clearly renders the corresponding equations irrelevant. Thus, the next result establishes a necessary and sufficient condition for the non-nullity of that summation.

**PROPOSITION 3.5.** *Let  $M, R \in \mathbb{N}$  such that  $M \geq R$  and  $M > 1$ ,  $m_1, \dots, m_P \in \{1, \dots, M\}$  and  $Q$  be defined as*

$$(3.12) \quad Q \triangleq \frac{M}{\gcd(M, m_1 + \dots + m_P - P)},$$

where  $\gcd(\cdot, \cdot)$  yields the greatest common divisor of its arguments. Then, we have

$$(3.13) \quad \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)} \neq 0$$

if and only if one of the following (mutually exclusive) conditions are met:

- (i)  $Q = 1$ ;
- (ii)  $Q$  does not divide  $R$ .

*Proof.* Defining  $v \triangleq (m_1 + \dots + m_P - P)/\gcd(M, m_1 + \dots + m_P - P) \in \mathbb{N}$  and using (3.12), we can obtain from (3.13) the equivalent relation

$$(3.14) \quad \sum_{r=1}^R \omega_Q^{-(r-1)v} \neq 0.$$

From (3.14), it is now evident that (i) implies (3.13), since  $\omega_1 = 1$ . On the other hand, if (ii) is true, then we necessarily have  $Q > 1$ . In this case, (3.14) corresponds to a sum of  $R$   $Q$ th roots of unity raised to a power  $v$  that is co-prime with  $Q$  by definition, which yields zero if and only if  $R = lQ$  for some positive integer  $l$ . Hence, (ii) also implies (3.13). Note that the last argument also establishes the only if part of the proof, since in this part we have to show that the sum in (3.13) vanishes if  $Q > 1$  and  $Q$  divides  $R$ . We observe also that, when these conditions hold, we have  $l \in \{1, \dots, d\}$  where  $d$  is the greatest integer that divides  $M$  and satisfies  $d < M$ , since  $Q$  divides  $M$  and  $R \leq M$ .  $\square$

**PROPOSITION 3.6.** *In the special case where  $M = R$  (i.e., the factors are square), condition (i) is necessary and sufficient for (3.13) of Proposition 3.5. Moreover, we can write it alternatively as*

$$(3.15) \quad \exists l \in \mathbb{N} \text{ such that } m_1 + \dots + m_P - P = lM.$$

*Proof.* The condition (ii) of Proposition 3.5 is immediately ruled out when  $M = R$ , since  $Q$  divides  $M$  by definition. The fact that the condition (i) and (3.15) are equivalent is trivial, and thus the proof is complete.  $\square$

We now focus on the characterization of the solutions to equations (3.11). First, we claim that there can be no “spurious” solution, in the sense that every solution to (3.11) provides a CPD of the desired form. This is assured by the following lemma.

**LEMMA 3.7.** *Let  $\mathcal{T}$  be a tensor such that the elements of  $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$  are given by (3.11), for some  $M, R$  satisfying  $M \geq R$  and some set of  $M$ -tuples  $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$ . Then, another set of  $M$ -tuples  $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$  denotes a solution to the equations of the form (3.11) if and only if it is associated with  $P$  circulant  $M \times R$  matrix factors constituting a CCPD of  $\mathcal{T}$ .*

*Proof.* The if part is clear from the analysis conducted to obtain the system of equations. For the only if part, we note that every solution  $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$  can be injectively associated with  $P$  circulant  $M \times R$  matrices  $\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)}$  whose circulant completions are given by  $\check{\mathbf{G}}^{(p)} = \mathbf{F}_M \text{diag}(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M}) \mathbf{F}_M^H$ . By construction, those matrices are such that  $\text{MDFT}\{\llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket\} = \mathcal{Y}$ , because the eigenvalues of  $\check{\mathbf{G}}^{(1)}, \dots, \check{\mathbf{G}}^{(P)}$  jointly satisfy the system of equations (3.11). Now, taking the IMDFT of both sides, we obtain  $\llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket = \mathcal{T}$ , as claimed.  $\square$

It should be noted that, in general, (3.11) admits infinitely many solutions. In light of the above result, this is of course expected, because of the ambiguities that are inherent to the CP model. In particular, if the factors  $\mathbf{C}^{(p)}$  in (3.5) satisfy Kruskal’s uniqueness condition, Theorem 2.1 tells us that the CPD is essentially unique, i.e., with only column scaling and permutation ambiguities on its factor matrices. This implies that the different solutions of (3.11) should also be related with each other accordingly. Our next result shows that such relation is rather simple.

**THEOREM 3.8.** *Let  $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$ , with  $P \geq 3$  and  $M > 1$ , be a hypercubic tensor of the form (3.5), whose factors  $\mathbf{C}^{(p)}$  are circulant  $M \times R$  matrices, with  $R > 1$ , that satisfy Kruskal’s condition  $\sum_{p=1}^P k_{\mathbf{C}^{(p)}} \geq 2R + P - 1$ . Let  $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$  denote the solution of (3.11) corresponding to the eigenvalues of the circulant completions of those factors and assume that  $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$  is another possible solution of (3.11). Then, there are  $P$  complex scalars  $\alpha_p$  satisfying  $\prod_{p=1}^P \alpha_p = 1$  and an integer  $r \in \{0, \dots, R-1\}$  such that, for all  $p \in \{1, \dots, P\}$  and  $m \in \{1, \dots, M\}$ , we have  $\mu_{p,m} = \alpha_p \omega_M^{-r(m-1)} \lambda_{p,m}$ .*

*Proof.* From Lemma 3.7, we have that the  $M$ -tuples  $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$  can be associated with circulant factors  $\mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)}$  such that  $\mathcal{T} = \llbracket \mathbf{G}^{(1)}, \dots, \mathbf{G}^{(P)} \rrbracket$ . More precisely,  $\{(\mu_{p,1}, \mu_{p,2}, \dots, \mu_{p,M})\}_{p=1}^P$  are the eigenvalues of  $\check{\mathbf{G}}^{(1)}, \dots, \check{\mathbf{G}}^{(P)}$ . But, since we assume that the factors satisfy Kruskal's uniqueness condition, we have

$$(3.16) \quad \mathbf{G}^{(p)} = \mathbf{C}^{(p)} \mathbf{\Pi} \mathbf{\Delta}_p$$

for some permutation matrix  $\mathbf{\Pi} \in \mathbb{R}^{R \times R}$  and some diagonal matrix  $\mathbf{\Delta}_p \in \mathbb{C}^{R \times R}$ , with the constraint  $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$ . Since by definition every column of  $\mathbf{C}^{(p)}$  is of the form  $\mathbf{\Pi}_M^r \mathbf{c}^{(p)}$ , where  $\mathbf{\Pi}_M$  is given by (3.2) and  $\mathbf{c}^{(p)}$  is the generating vector of  $\mathbf{C}^{(p)}$ , we have from (3.16) that there exists  $r \in \{0, \dots, R-1\}$  such that

$$(3.17) \quad \mathbf{g}^{(p)} = \alpha_p \mathbf{\Pi}_M^r \mathbf{c}^{(p)},$$

where  $\mathbf{g}^{(p)}$  is the generating vector of  $\mathbf{G}^{(p)}$  and  $\alpha_p = [\mathbf{\Delta}_p]_{1,1}$ . Due to the circulant structure of  $\mathbf{G}^{(p)}$ , we can also write

$$(3.18) \quad \mathbf{G}^{(p)} = \alpha_p \mathbf{\Pi}_M^r \mathbf{C}^{(p)}.$$

Now, substituting (3.4) in (3.17) and premultiplying both sides by  $\sqrt{M} \mathbf{F}_M^H$ , we obtain

$$\sqrt{M} \mathbf{F}_M^H \mathbf{g}^{(p)} = \alpha_p \sqrt{M} \mathbf{F}_M^H \mathbf{F}_M \text{diag} \left( 1, \omega_M^{-r}, \dots, \omega_M^{-r(M-1)} \right) \mathbf{F}_M^H \mathbf{c}^{(p)}.$$

But, since  $\mathbf{G}^{(p)}$  and  $\mathbf{C}^{(p)}$  are circulant, we have  $\sqrt{M} \mathbf{F}_M^H \mathbf{g}^{(p)} = \boldsymbol{\mu}_p$  and  $\sqrt{M} \mathbf{F}_M^H \mathbf{c}^{(p)} = \boldsymbol{\lambda}_p$ , where  $\boldsymbol{\mu}_p = [\mu_{p,1} \dots \mu_{p,M}]^T$  and  $\boldsymbol{\lambda}_p = [\lambda_{p,1} \dots \lambda_{p,M}]^T$ . Using this property and recalling that  $\mathbf{F}_M$  is unitary, we can rewrite the above equation as  $\boldsymbol{\mu}_p = \alpha_p \text{diag} \left( 1, \omega_M^{-r}, \dots, \omega_M^{-r(M-1)} \right) \boldsymbol{\lambda}_p$ , which, in scalar form, corresponds to  $\mu_{p,m} = \alpha_p \omega_M^{-r(m-1)} \lambda_{p,m}$ . To complete the proof, we observe that the constraint  $\mathbf{\Delta}_1 \dots \mathbf{\Delta}_P = \mathbf{I}_R$  implies  $\prod_{p=1}^P \alpha_p = 1$  and that the constant  $r \in \{0, \dots, R-1\}$  is the same for all  $p \in \{1, \dots, P\}$ , since the same permutation  $\mathbf{\Pi}$  applies to all factors in (3.16).  $\square$

**COROLLARY 3.9.** *If  $M = R$ , then  $\mathbf{\Pi} = \mathbf{\Pi}_M^r$ , where  $r$  is the same integer as in (3.18), and  $\mathbf{\Delta}_p = \alpha_p \mathbf{I}_R$  for all  $p$ .*

*Proof.* From Kruskal's condition,  $k_{\mathbf{C}^{(q)}} > 1$  must hold for at least some  $q \in \{1, \dots, P\}$ . Let us consider two distinct columns of  $\mathbf{G}^{(q)}$ , which can be written as  $\mathbf{g}_k^{(q)} = \alpha_{q,k} \mathbf{\Pi}_M^m \mathbf{c}^{(q)}$  and  $\mathbf{g}_l^{(q)} = \alpha_{q,l} \mathbf{\Pi}_M^n \mathbf{c}^{(q)}$  for some  $m, n \in \{0, \dots, M-1\}$ , with  $l > k$  and  $\alpha_{q,k}, \alpha_{q,l} \neq 0$ . As  $\mathbf{G}^{(q)}$  is circulant, we also have  $\mathbf{g}_l^{(q)} = \mathbf{\Pi}_M^{l-k} \mathbf{g}_k^{(q)}$ , which implies  $\mathbf{\Pi}_M^n \mathbf{c}^{(q)} = \beta \mathbf{\Pi}_M^{l-k+m} \mathbf{c}^{(q)}$ , with  $\beta = \alpha_{q,k} / \alpha_{q,l}$ . Now, if  $\beta \neq 1$ , then  $k_{\mathbf{C}^{(q)}} \leq 1$ , contradicting our hypothesis. Using the same reasoning for all couples of columns of  $\mathbf{G}^{(q)}$ , we deduce that  $\mathbf{\Delta}_q = \alpha_q \mathbf{I}_R$ . Moreover, since all circulant matrices commute, we have from (3.18) that  $\mathbf{G}^{(q)} = \alpha_q \mathbf{C}^{(q)} \mathbf{\Pi}_M^r$ . Substituting these results in (3.16) with  $p = q$ , it follows that  $\mathbf{\Pi} = \mathbf{\Pi}_M^r$ , which in turn implies  $\mathbf{\Delta}_p = \alpha_p \mathbf{I}_R$  for all  $p$ .  $\square$

**REMARK 3.10.** *As seen above, when  $M = R$  the permutation matrix of (3.16) must be circulant. Conversely, all the  $M$  distinct circulant permutations of the original factors yield an equivalent CCPD. In other words, if  $\{(\lambda_{p,1}, \lambda_{p,2}, \dots, \lambda_{p,M})\}_{p=1}^P$  is a solution for (3.11), then so is any set  $\{(\omega_M^0 \lambda_{p,1}, \omega_M^{-r} \lambda_{p,2}, \dots, \omega_M^{-r(M-1)} \lambda_{p,M})\}_{p=1}^P$  with  $r \in \{0, \dots, M-1\}$ . Indeed, from Proposition 3.6 we know that, when  $M = R$ , every nonzero equation of the form (3.11) is associated with indices  $m_1, \dots, m_P$  that satisfy  $m_1 + \dots + m_P - P = lM$  for some  $l \in \mathbb{N}$ , and hence  $\prod_{p=1}^P \lambda_{p,m_p} = \omega_M^{-r(m_1 + \dots + m_P - P)} \prod_{p=1}^P \lambda_{p,m_p} = \prod_{p=1}^P \omega_M^{-r(m_p - 1)} \lambda_{p,m_p}$ .*

TABLE 1  
Summary of the AS method.

**Inputs:** Hypercubic tensor  $\mathcal{T} \in \mathbb{C}^{M^{(P)}}$  to be decomposed and the rank  $R$  of the CCPD.

**Outputs:** Circulant  $M \times R$  complex matrix factors  $\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)}$  (or  $\mathbf{C}$ , in the symmetric case).

1. Compute  $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$ .
2. Find the relevant equations using Proposition 3.5 (or Proposition 3.6, if  $M = R$ ).
3. Solve the relevant equations (3.11) (resp., (3.20)) for  $\lambda_{p,m}$  (resp.,  $\lambda_m$ ).
4. Reconstruct the factors  $\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(P)}$  (resp.,  $\mathbf{C}$ ) by resorting to property (3.3).

**3.2. Symmetric CCPD.** The symmetric case can be derived from the previous subsection by considering that  $\mathbf{C}^{(1)} = \dots = \mathbf{C}^{(P)} = \mathbf{C} \in \mathbb{C}^{M \times R}$ , which yields

$$(3.19) \quad \mathcal{T} = \llbracket \mathbf{C}, \dots, \mathbf{C} \rrbracket \in \mathbb{C}^{M^{(P)}}.$$

As we shall see, this simplifies the set of solutions of equations (3.11). First, let us rewrite those equations as

$$(3.20) \quad y_{m_1, \dots, m_P} = \left( \frac{1}{\sqrt{M}} \right)^P \prod_{p=1}^P \lambda_{m_p} \sum_{r=1}^R \omega_M^{-(r-1)(m_1 + \dots + m_P - P)},$$

where the first subscript of the eigenvalues has been dropped, since they all refer now to the same circulant completion,  $\tilde{\mathbf{C}}$ . This particularization clearly renders some of the equations (3.20) redundant, because the MDFT of a symmetric tensor is also symmetric. Moreover, assuming that Kruskal's condition holds, i.e., that  $k_{\mathbf{C}} \geq (2R - 1)/P + 1$ , we can show that the set of solutions of (3.20) is finite, having at most  $PM$  elements. Once again, this is demonstrated by exploiting the link between these solutions and the set of factors which can yield a symmetric CCPD of  $\mathcal{T}$ , as follows.

**THEOREM 3.11.** *Let  $\mathcal{T}$  be given by (3.19), where  $\mathbf{C} \in \mathbb{C}^{M \times R}$ , with  $M \geq R > 1$ , is circulant. If  $k_{\mathbf{C}} \geq (2R - 1)/P + 1$ , then the equations (3.20) admit at most  $PM$  different solutions.*

*Proof.* The proof is similar to that of Theorem 3.8, with the difference that we must now have  $\alpha_1 = \dots = \alpha_P = \alpha$ , since the CCPD (3.19) is symmetric. Due to the constraint  $\Delta_1 \dots \Delta_P = \mathbf{I}_R$ , the only possible distinct values for  $\alpha$  are the  $P$ th roots of unity, i.e.,  $\alpha = \omega_P^p$  for  $p \in \{0, \dots, P-1\}$ . Combining this with the existence of only  $M$  distinct values for the integer  $r$  in (3.17), we are left with at most  $PM$  distinct solutions for (3.20).  $\square$

**REMARK 3.12.** *From the result of Corollary 3.9 we conclude that, if a symmetric CCPD with square factors is essentially unique, then (3.20) admits exactly  $PM$  solutions, since every combination of one of the  $P$  distinct matrices  $\omega_P^p \mathbf{I}_R$  with one of the  $M$  distinct matrices  $\Pi_M^r$  yields an equivalent symmetric CCPD.*

**3.3. The AS method.** As we have shown, *any* solution of the equations (3.11) provides a set of  $P$  factors which jointly fulfill (3.5), as desired. Therefore, we conclude that the resolution of such equations provides an *exact* solution for the CCPD in the noiseless case. In the presence of noise, evidently, only an approximate solution of (3.11) can be sought.

Table 1 summarizes the approach proposed in this paper, which we shall refer to as an algebraic solution (AS) for the CCPD. With regard to the first step, we point out that  $\mathcal{Y}$  can be computed in an efficient way, through a multidimensional FFT

algorithm [13]. The second step can be performed by verifying for all combinations of indices  $m_1, \dots, m_P \in \{1, \dots, M\}$  which of them satisfy one of the conditions of Proposition 3.5, if  $M > R$ , or the condition of Proposition 3.6, if  $M = R$ . It is evident that, since for each combination of indices these conditions depend only on  $P$ ,  $R$  and  $M$ , their assessment can be done *a priori* and reused for several different hypercubic tensors of same order, rank and dimension. Step 3 is the core of the method, being discussed below. Finally, step 4 consists in simply reconstructing the desired factors from the estimated eigenvalues of their circulant completions.

The resolution of the set of monomial equations (i.e., step 3 of Table 1) is a central matter in our approach. However, discussing computational methods for the resolution of polynomial systems, which have been employed to solve practical problems (see, e.g., [24] and references therein), is outside the scope of this paper. Here, we consider only some simple examples of the systems (3.11) and (3.20), where the solution of the monomial equations can be easily derived in a direct, non-optimal, manner. This derivation is done in §4, which provides an illustration of our theoretical results. Moreover, the derived procedures will be used for numerical evaluation purposes in §5. This kind of *ad hoc* approach, which we shall refer to as AAS (*ad-hoc* algebraic solution), is of very low computational cost and can always be derived, as long as some assumptions about the non-nullity of certain eigenvalues hold. These assumptions hold generically for hypercubic tensors  $\mathcal{T}$  of the form (3.5) having tall or square factors, and therefore little generality is lost. Although the accuracy of the AAS estimate is degraded in the presence of noise, it can be effectively and efficiently refined with iterative algorithms, as will be illustrated by the experimental results presented in §5. Indeed, these results show that iterative algorithms initialized with the AAS estimate can avoid local minima and converge relatively fast, performing substantially better than when randomly initialized.

**3.4. Comparison with existing estimation algorithms for structured CPDs.** To the best of our knowledge, no previous work has developed an approach capable of exploiting the structure of a CCPD completely. However, the estimation of banded circulant factors of a CPD can be done with the use of some existing algorithms, which are briefly described in the following.

In [21], a non-iterative algorithm which decomposes a third-order tensor having one banded circulant factor is proposed, whose extension to higher-order cases is trivial. Basically, a basis for the column space of the structured factor is obtained from the SVD of a matrix unfolding of the tensor, and then a matrix with the desired structure is found by solving a system of linear equations constructed from this basis. Then, the other factors are estimated by computing  $R$  best rank-one approximations of  $(P-1)$ th-order tensors,<sup>2</sup> where  $P$  is the order of the tensor of interest. A drawback of this method is that it cannot simultaneously take into account the structure of multiple factors. Furthermore, the circulant factor must be of a particular banded form which limits its rank.

A similar strategy proposed in [10] and then further developed in [31] is able to jointly estimate banded structured factors of a CPD non-iteratively. It applies to both Toeplitz and Hankel-structured factors. However, at most structured  $P-1$  factors can be found by solving the resulting system of linear equations, because the matrix unfolding whose SVD is computed induces a partition of the factors into two sets, of

---

<sup>2</sup>It should be noted that the best rank-one approximation of a tensor is a well posed problem, as opposed to finding the best rank- $R$  approximation [12], but requires the use of an iterative algorithm if this tensor has order greater than two.

which only one has the structure taken into account when constructing such system. Moreover, when jointly estimating  $P_s > 2$  structured factors, an iterative algorithm must be used to solve a rank-one approximation problem of a  $P_s$ th-order tensor.

We conclude that, in comparison with previous approaches, the main advantages of the AS are the absence of additional constraints over the factors, which need not be banded, and its ability to take into account the complete structure of the CCPD. In addition, its specialization to the symmetric case is straightforward, while this constraint would have to be imposed *a posteriori* in [21, 31]. On the other hand, the algorithms of [21, 31] can directly handle CPDs consisting not only of structured factors, and are also attractive for relying on numerically robust linear algebra techniques. A quantitative comparison involving the AS approach and these algorithms will be presented in §5, by means of numerical simulations.

#### 4. Illustrative examples.

**EXAMPLE 4.1.** Let  $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)} \in \mathbb{C}^{3 \times 2}$  be circulant matrices with generating vectors  $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}$  and  $\mathbf{c}^{(3)}$ , respectively, and consider the CCPD  $\mathcal{T} = [\![\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)}]\!]$ . We can obtain from the nonzero elements of  $\mathcal{Y} = \text{MDFT}\{\mathcal{T}\}$  27 equations in the 9 unknowns of interest, which are the eigenvalues of  $\check{\mathbf{C}}^{(1)}, \check{\mathbf{C}}^{(2)}$  and  $\check{\mathbf{C}}^{(3)}$ . In particular,

$$(4.1) \quad \lambda_{1,1} \lambda_{2,1} \lambda_{3,1} = \frac{\tilde{y}_{1,1,1}}{2},$$

where  $\tilde{y}_{m_1, m_2, m_3} \triangleq 3\sqrt{3} y_{m_1, m_2, m_3}$ . We know also that an infinite number of solutions exist, which is a consequence of the scaling ambiguity. Yet, since  $\lambda_p/\sqrt{3}$  is the DFT of  $\mathbf{c}^{(p)}$ , under the assumption that  $\lambda_{1,1} \lambda_{2,1} \lambda_{3,1} \neq 0$  we can conclude that the value of each  $\lambda_{p,1}$  determines the scaling of its associated factor  $\mathbf{C}^{(p)}$ . In this case, we can eliminate that ambiguity by imposing the values of all but one  $\lambda_{p,1}$ . For instance, if we choose  $\lambda_{2,1} = \lambda_{3,1} = 1$ , then (4.1) can be directly solved for  $\lambda_{1,1}$ . Next, we can use, for example, the equations

$$\begin{aligned} \lambda_{1,1} \lambda_{2,1} \lambda_{3,2} &= \frac{\tilde{y}_{1,1,2}}{1 + \omega_3^{-1}} \rightarrow \lambda_{3,2}, & \lambda_{1,2} \lambda_{2,1} \lambda_{3,1} &= \frac{\tilde{y}_{2,1,1}}{1 + \omega_3^{-1}} \rightarrow \lambda_{1,2}, \\ \lambda_{1,1} \lambda_{2,3} \lambda_{3,1} &= \frac{\tilde{y}_{1,3,1}}{1 + \omega_3^{-2}} \rightarrow \lambda_{2,3}, & \lambda_{1,1} \lambda_{2,2} \lambda_{3,1} &= \frac{\tilde{y}_{1,2,1}}{1 + \omega_3^{-1}} \rightarrow \lambda_{2,2}, \\ \lambda_{1,1} \lambda_{2,1} \lambda_{3,3} &= \frac{\tilde{y}_{1,1,3}}{1 + \omega_3^{-2}} \rightarrow \lambda_{3,3}, & \lambda_{1,3} \lambda_{2,1} \lambda_{3,1} &= \frac{\tilde{y}_{3,1,1}}{1 + \omega_3^{-2}} \rightarrow \lambda_{1,3} \end{aligned}$$

to find the values of the other eigenvalues, as indicated. With those eigenvalues at hand, the reconstruction of the desired factors is trivial via relation (3.3). We emphasize that other choices of equations are possible for obtaining a solution and that the procedure described above is by no means guaranteed to be optimal in any sense.

After resolving the scaling indeterminacy, obtaining the corresponding solution is straightforward in the Example 4.1, provided that  $\lambda_{1,1} \lambda_{2,1} \lambda_{3,1} \neq 0$ . This simple approach can be employed in general, but other instances may involve more complicated procedures and may require the non-nullity of other eigenvalues. Nonetheless, when applicable, the resulting procedure is computationally inexpensive, and thus may be used to initialize more complicated methods, as shown in the next section.

**EXAMPLE 4.2.** For the symmetric case where  $P = M = R = 3$ , defining  $\tilde{y}_{m_1, m_2, m_3} \triangleq \sqrt{3} y_{m_1, m_2, m_3}$  and disregarding redundant equations due to symmetries, we have the following equations:  $\lambda_1^3 = \tilde{y}_{1,1,1}$ ,  $\lambda_2^3 = \tilde{y}_{2,2,2}$ ,  $\lambda_3^3 = \tilde{y}_{3,3,3}$  and  $\lambda_1 \lambda_2 \lambda_3 =$



TABLE 2  
Summary of the evaluated methods.

Acronym	Algorithm	Initialization	Acronym	Algorithm	Initialization
AAS	<i>ad-hoc</i> AS	—	CPTOEP	[31]	—
ALS	ALS	random	CPTOEP-ALS	ALS	CPTOEP
AAS-ALS	ALS	AAS	CPTOEP-CALS	CALS	CPTOEP
CALS	CALS	random	CPTOEP-LMAS	LM	CPTOEP
AAS-CALS	CALS	AAS	TPARAFAC1	[21]	—
LMAS	LM	random	TPARAFAC1-ALS	ALS	TPARAFAC1
AAS-LMAS	LM	AAS	TPARAFAC1-LMAS	LM	TPARAFAC1

$\tilde{y}_{1,2,3}$ . One can promptly see that we now have three uncoupled equations for calculating the three unknowns  $\lambda_1, \lambda_2, \lambda_3$ . However, the coupled equation in  $\lambda_1 \lambda_2 \lambda_3$  constrains the possible solutions, so that there can be at most 9 instead of 27 solutions, which correspond precisely to the maximum number of PM solutions stated by Theorem 3.11.

REMARK 4.3. If it is known a priori that  $\mathbf{C} \in \mathbb{R}^{M \times R}$  and  $\lambda_1 \neq 0$ , we can choose a real solution for the equation involving  $\lambda_1^P$ , which is guaranteed to exist due to (3.10). This allows suppressing the scaling ambiguity whilst obtaining a real factor  $\mathbf{C}$ .

**5. Numerical simulations.** In order to evaluate the proposed approach and compare it with existing ones, we present now two Monte Carlo simulation scenarios. In the first one, we measure the performance of several estimation methods when used to compute non-symmetric CCPDs whose factors are randomly generated. The second scenario illustrates the practical applicability of our approach by comparing its performance with those of existing ones in a homogeneous Wiener-Hammerstein identification problem, where a CPD structurally close to a symmetric CCPD arises.

Table 2 lists the methods considered in our comparison. The acronym LMAS stands for “Levenberg-Marquardt algorithm applied to the AS”, since the LM scheme is employed to solve the monomial equations. The algorithms referred to as CPTOEP and TPARAFAC1 are those proposed in [31] and [21], respectively. Note that we have evaluated the performance of iterative algorithms with random initialization and also with initialization given by the non-iterative ones, yielding combinations which are denoted by the hyphen-separated acronyms. For instance, AAS-LMAS consists in employing the LMAS algorithm initialized by the output of AAS.

First of all, however, we formulate a modified ALS algorithm for third order tensors which constrains the factors to be circulant. Since we are interested in factors with such structure, this version, which we shall call CALS (for circulant-constrained ALS), is a more appropriate reference for our evaluation than the standard ALS.

**5.1. CALS algorithm.** Let  $\mathcal{C}^{M \times R} \subset \mathbb{C}^{M \times R}$  be the subspace of circulant  $M \times R$  matrices, and consider the basis  $\{\mathbf{E}_m\}_{m=1}^M$ , where the generating vector of  $\mathbf{E}_m$  is the canonical basis vector  $\mathbf{e}_m$  of  $\mathbb{C}^M$ . Clearly, any matrix  $\mathbf{C}^{(p)} \in \mathcal{C}^{M \times R}$  with generating vector  $\mathbf{c}^{(p)} = [c_1^{(p)} \ \dots \ c_M^{(p)}]^T$  can be written as

$$(5.1) \quad \mathbf{C}^{(p)} = \sum_{m=1}^M c_m^{(p)} \mathbf{E}_m = \text{unvec}_R \left( \mathbf{E} \mathbf{c}^{(p)} \right),$$

where  $\mathbf{E} = [\text{vec}(\mathbf{E}_1) \ \dots \ \text{vec}(\mathbf{E}_M)] \in \mathbb{C}^{RM \times M}$  and the  $\text{unvec}_R$  operator is such that  $\text{unvec}_R(\mathbf{a}) = [\mathbf{a}_1 \ \dots \ \mathbf{a}_R] \in \mathbb{C}^{N \times R}$  for every vector  $\mathbf{a} = [\mathbf{a}_1^T \ \dots \ \mathbf{a}_R^T]^T \in \mathbb{C}^{RN}$  with  $\mathbf{a}_1, \dots, \mathbf{a}_R \in \mathbb{C}^N$ . Hence, if a third-order tensor  $\mathcal{T}$  is given by (3.5), from (2.10) and



TABLE 3  
The CALS algorithm.

**Inputs:**  $\mathcal{T} \in \mathbb{C}^{M \times M \times M}$ , rank  $R$  of the CCPD and initial generating vectors  $\mathbf{c}_0^{(1)}, \mathbf{c}_0^{(2)}, \mathbf{c}_0^{(3)}$ .

**Outputs:** Estimated generating vectors  $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)}$ .

$i = 1$

**repeat**

$$\begin{aligned} \mathbf{c}_i^{(1)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_1) \text{vec} \left\{ \left[ \left( \text{circ}_R(\mathbf{c}_{i-1}^{(3)}) \diamond \text{circ}_R(\mathbf{c}_{i-1}^{(2)}) \right)^T \right]^\dagger \right\} \\ \mathbf{c}_i^{(2)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_2) \text{vec} \left\{ \left[ \left( \text{circ}_R(\mathbf{c}_i^{(1)}) \diamond \text{circ}_R(\mathbf{c}_{i-1}^{(3)}) \right)^T \right]^\dagger \right\} \\ \mathbf{c}_i^{(3)} &= \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_3) \text{vec} \left\{ \left[ \left( \text{circ}_R(\mathbf{c}_i^{(2)}) \diamond \text{circ}_R(\mathbf{c}_i^{(1)}) \right)^T \right]^\dagger \right\} \end{aligned}$$

$i = i + 1$

**until** convergence

(5.1) we deduce that its unfolding  $\mathbf{T}_1$  is given by

$$\mathbf{T}_1 = \mathbf{C}^{(1)} \left( \mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T = \text{unvec}_R \left( \mathbf{E} \mathbf{c}^{(1)} \right) \left( \mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T.$$

Now, if we assume  $\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)}$  is full-column rank (which always holds for full column-rank  $\mathbf{C}^{(2)}$  and  $\mathbf{C}^{(3)}$ ), we have  $\text{unvec}_R(\mathbf{E} \mathbf{c}^{(1)}) = \mathbf{T}_1 \left[ (\mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)})^T \right]^\dagger$ . Then, applying  $\text{vec}$  on both sides and multiplying them by  $\mathbf{E}^\dagger = \frac{1}{R} \mathbf{E}^T$  we obtain

$$(5.2) \quad \mathbf{c}^{(1)} = \frac{1}{R} \mathbf{E}^T (\mathbf{I}_R \otimes \mathbf{T}_1) \text{vec} \left\{ \left[ \left( \mathbf{C}^{(3)} \diamond \mathbf{C}^{(2)} \right)^T \right]^\dagger \right\},$$

where we have used the property  $\text{vec}(\mathbf{ABC}^T) = (\mathbf{C} \otimes \mathbf{A}) \text{vec}(\mathbf{B})$  with  $\mathbf{C} = \mathbf{I}_R$ . Expression (5.2) provides an update rule for the generating vector  $\mathbf{c}^{(1)}$ , instead of a general factor matrix. This allows us to exploit the structure of  $\mathbf{C}^{(1)}$  for obtaining better-estimated parameters. The resulting algorithm is summarized in Table 3.

**5.2. General CCPD.** Initially, we independently draw the entries of the  $P = 3$  generating vectors  $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \mathbf{c}^{(3)} \in \mathbb{C}^M$  from a standard Gaussian distribution, setting  $[\mathbf{c}^{(p)}]_1 = 1$  for all  $p$ , though, to avoid the scaling ambiguity. Next,  $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)} \in \mathbb{C}^{M \times R}$  are constructed according to (3.1), and  $\mathcal{T}$  is obtained with (3.5). The data tensor is then generated via  $\mathcal{D} = \mathcal{T} + \sigma_N \mathcal{N}$ , where  $\mathcal{N} \in \mathbb{C}^{M \times M \times M}$  is a noise tensor, whose elements are independently drawn from a standard Gaussian distribution and normalized to ensure  $\|\mathcal{N}\|_F = 1$ , and  $\sigma_N$  is a positive parameter used to impose the desired level of SNR, which is defined as  $\text{SNR} = 10 \log_{10} \|\mathcal{T}\|_F^2 / \sigma_N^2$  dB. Once the data tensor is available, we employ each estimation method to obtain estimates  $\hat{\mathbf{C}}^{(1)}, \hat{\mathbf{C}}^{(2)}$  and  $\hat{\mathbf{C}}^{(3)}$  such that  $\hat{\mathcal{T}} = \llbracket \hat{\mathbf{C}}^{(1)}, \hat{\mathbf{C}}^{(2)}, \hat{\mathbf{C}}^{(3)} \rrbracket$  yields the reconstructed tensor. The resulting reconstruction error is then calculated with  $\text{NMSE}_{\mathcal{T}} = \|\mathcal{D} - \hat{\mathcal{T}}\|_F^2 / \|\mathcal{D}\|_F^2$ . We compute also the average normalized square error over the estimated generating vectors, given by  $\varepsilon = (1/3) \sum_{p=1}^3 \|\mathbf{c}^{(p)} - \hat{\mathbf{c}}^{(p)}\|_2^2 \times \|\mathbf{c}^{(p)}\|_2^{-2}$ . In order to circumvent the permutation ambiguity, the estimates  $\hat{\mathbf{c}}^{(p)}$  are chosen as the corresponding columns of the factors which, after being normalized so that  $[\hat{\mathbf{c}}^{(p)}]_1 = 1$ , jointly minimize  $\varepsilon$ .

We compared all the methods of Table 2, except for those involving the algorithm TPARAFAC1, since it cannot estimate non-banded factors. Although CPTOEP is

also intended to compute banded factors, it can be easily applied to jointly estimate non-banded circulant factors, by considering an appropriate basis for each structured factor when constructing the system of equations. With this strategy, we first perform the joint estimation of  $\hat{\mathbf{C}}^{(1)}$  and  $\hat{\mathbf{C}}^{(2)}$ , then we estimate  $\hat{\mathbf{c}}^{(3)}$  in the least-squares sense from the recovered third factor (see [31, Sec. 6.2]) and construct  $\hat{\mathbf{C}}^{(3)} = \text{circ}_R(\hat{\mathbf{c}}^{(3)})$ . However, while this procedure succeeds in the tall case, it fails when the factors are square, which seems to be due to the relaxation adopted to solve [31, eq. (28)].

As the estimates of ALS are not necessarily circulant, we first choose each  $\hat{\mathbf{c}}^{(p)}$  as described above and then construct  $\hat{\mathbf{C}}^{(p)} = \text{circ}_R(\hat{\mathbf{c}}^{(p)})$ , imposing the structure *a posteriori*. Concerning the use of the LM scheme to solve the monomial equations, we note that the scaling indeterminacy implies the existence of degenerate critical points in the parameters space, causing a near-singularity of the Jacobian matrix in some iterations. The LMAS method avoids this by imposing  $\lambda_{2,1} = \lambda_{3,1} = 1$ , similarly to what was done in Example 4.1. Note that this does not imply any loss of generality as long as  $\lambda_{2,1}\lambda_{3,1} \neq 0$ , which holds with probability one in our simulations. The adopted cost function is given by  $J(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3) = \|\mathcal{R}\|_F^2 / \|\mathcal{Y}\|_F^2$ , where  $\mathcal{R} = \mathcal{R}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3) \in \mathbb{C}^{M^{(P)}}$  is a residue tensor whose elements satisfy

$$[\mathcal{R}]_{m_1, \dots, m_P} = [\mathcal{Y}]_{m_1, \dots, m_P} - M^{-\frac{3}{2}} \lambda_{1,m_1} \lambda_{2,m_2} \lambda_{3,m_3} \sum_{r=1}^R \omega_M^{-(r-1)(m_1+m_2+m_3-3)}.$$

For a fair comparison, all the iterative algorithms were assumed to converge whenever  $(\|\mathcal{D} - \hat{\mathcal{T}}_i\|_F^2 - \|\mathcal{D} - \hat{\mathcal{T}}_{i-1}\|_F^2) / \|\mathcal{D}\|_F^2 < 10^{-6}$ , where  $\hat{\mathcal{T}}_i$  denotes the estimated reconstructed tensor at iteration  $i$ , or when the maximum number of 2000 iterations was reached.

The simulations were performed for two configurations: (a)  $M = 3$ ,  $R = 2$  and (b)  $M = R = 3$ . In each configuration, the above described procedure was repeated for 1000 Monte Carlo runs, with the SNR varying in  $[20, 80]$  dB, and the average  $\varepsilon$  and  $\text{NMSE}_{\mathcal{T}}$  were calculated. Before computing these averages, we discard, for each method and each SNR level, the 1% of realizations with the highest  $\varepsilon$ , to avoid misleading conclusions brought by the occurrence of a few outstandingly poor estimates which were observed in practice. Table 4 summarizes the results. Note that, as CPTOEP was not able to estimate square non-banded factors, it was not included in (b). To facilitate the comparison, we underline the lowest error(s) at each SNR level. We can see that the average errors of AAS steadily drop as the SNR is increased. The same is not true for the randomly initialized iterative methods, which perform poorly due to frequent convergence to local minima. When initialized with the AAS, nonetheless, they are able to refine it and attain lower average error levels, which indicates that local minima are avoided. This behavior is also observed in (a) for the methods initialized with CPTOEP, which provides a better initial estimate than AAS, but at the expense of a higher computational cost, because of the simplicity of the AAS procedure. Overall, AAS-LMAS, AAS-CALS and CPTOEP-CALS attain the best performances in (a), while AAS-LMAS is the most accurate in (b).

**5.3. Homogeneous Wiener-Hammerstein system identification.** Consider the identification of a discrete-time Wiener-Hammerstein system consisting of a non-linear homogeneous block  $F(x) = x^P$  preceded by a linear time-invariant (LTI) FIR filter  $\mathbf{w} \in \mathbb{C}^{L_w}$  and succeeded by another LTI FIR filter  $\mathbf{h} \in \mathbb{C}^{L_h}$ . It is well known

that such system is equivalent to a  $P$ th-order homogeneous Volterra filter of the form

$$y(n) = \sum_{m_1=1}^M \cdots \sum_{m_P=1}^M [\mathcal{K}]_{m_1, \dots, m_P} \prod_{p=1}^P u(n - m_p),$$

where  $u(n)$  and  $y(n)$  are, respectively, its input and output, having a  $P$ th-order symmetric Volterra kernel  $\mathcal{K} \in \mathbb{C}^{M^{(P)}}$  with memory  $M = L_w + L_h - 1$  such that [21]

$$[\mathcal{K}]_{m_1, \dots, m_P} = \sum_{l=l_0}^L [\mathbf{h}]_l \prod_{p=1}^P [\mathbf{w}]_{m_p - l + 1},$$

with  $L = \min\{L_h, m_1, \dots, m_P\}$  and  $l_0 = \max\{1, m_1 - L_w + 1, \dots, m_P - L_w + 1\}$ . In the identification problem, one is interested in estimating  $\mathbf{w}$  and  $\mathbf{h}$  from a set of input/output samples. Due to the above described equivalence, this can be done by first estimating the Volterra kernel  $\mathcal{K}$  from input/output data and then computing its CPD, which satisfies [21]

$$(5.3) \quad \mathcal{K} = [\mathbf{W} \text{diag}(\mathbf{h}), \dots, \mathbf{W}, \mathbf{W}],$$

where  $\mathbf{W} = \text{circ}_{L_h}([\mathbf{w}^T \ \mathbf{0}^T]^T) \in \mathbb{C}^{M \times L_h}$ . Note that the choice of which factor is post-multiplied by  $\text{diag}(\mathbf{h})$  is irrelevant, due to the scaling ambiguity. From the discussion of Remark 3.4, we have that the elements of  $\mathcal{Y} = \text{MDFT}\{\mathcal{K}\}$  are given by

$$(5.4) \quad y_{m_1, \dots, m_P} = \left( \frac{1}{\sqrt{M}} \right)^P \prod_{p=1}^P \lambda_{m_p} \sum_{l=1}^{L_h} [\mathbf{h}]_l \omega_M^{-(l-1)(m_1 + \dots + m_P - P)},$$

TABLE 4  
Simulation results for the general CCPD scenario.

		(a) $M = 3, R = 2$									
		$\varepsilon$ (dB)					NMSE $_{\mathcal{T}}$ (dB)				
SNR (dB)	Method	20	30	40	50	60	20	30	40	50	60
	ALS	-22.6	-23.8	-31.6	-25.4	-27.1	-17.9	-20.0	-28.2	-21.3	-23.0
	CALS	-13.4	-11.7	-12.3	-14.6	-13.5	-14.1	-13.2	-13.1	-15.6	-14.0
	LMAS	-26.1	-29.5	-30.8	-29.1	-29.4	-24.2	-27.9	-29.4	-26.3	-26.7
	AAS	-8.9	-17.7	-27.7	-37.6	-47.6	18.2	-6.7	-22.2	-32.4	-42.3
	AAS-ALS	-22.8	-33.0	-42.9	-51.6	-59.3	-18.1	-28.2	-38.2	-46.9	-54.7
	AAS-CALS	-27.4	-37.5	-47.5	-57.4	-67.2	-25.6	-35.6	-45.6	-55.6	-65.4
	AAS-LMAS	<u>-27.6</u>	<u>-37.7</u>	<u>-47.8</u>	<u>-57.7</u>	<u>-67.7</u>	<u>-25.8</u>	<u>-35.8</u>	<u>-45.8</u>	<u>-55.8</u>	<u>-65.8</u>
	CPTOEP	-25.5	-35.6	-45.6	-55.6	-65.6	-23.9	-33.9	-44.0	-54.0	-64.0
	CPTOEP-ALS	-22.9	-33.3	-43.9	-54.2	-64.2	-18.2	-28.6	-39.2	-49.6	-59.6
	CPTOEP-CALS	-27.4	-37.5	-47.5	-57.4	-67.4	-25.6	-35.6	-45.6	-55.6	-65.6
	CPTOEP-LMAS	-27.5	-36.1	-39.7	-39.5	-39.6	-25.7	-34.4	-38.2	-38.7	-38.8
		(b) $M = R = 3$									
		$\varepsilon$ (dB)					NMSE $_{\mathcal{T}}$ (dB)				
SNR (dB)	Method	20	30	40	50	60	20	30	40	50	60
	ALS	-14.7	-16.6	-16.0	-16.7	-16.6	-9.2	-11.4	-11.8	-11.6	-11.8
	CALS	<u>-22.7</u>	-31.7	-39.5	-34.5	-40.1	<u>-25.8</u>	-35.8	-45.6	-52.6	-56.0
	LMAS	-13.1	-13.1	-12.4	-13.8	-12.4	-20.6	-21.9	-21.0	-21.6	-21.1
	AAS	-10.1	-18.1	-27.8	-37.7	-47.6	9.2	-12.2	-23.4	-33.6	-43.8
	AAS-ALS	-17.4	-28.2	-38.0	-45.7	-53.3	-12.8	-23.6	-33.5	-41.7	-50.1
	AAS-CALS	-22.2	-32.5	-40.5	-46.5	-53.8	<u>-25.8</u>	-35.8	-45.6	-54.6	-62.9
	AAS-LMAS	-21.0	<u>-32.9</u>	<u>-43.1</u>	<u>-53.0</u>	<u>-63.1</u>	<u>-25.6</u>	<u>-35.9</u>	<u>-45.9</u>	<u>-55.8</u>	<u>-65.9</u>

where  $\lambda_m$  is the  $m$ th component of  $\boldsymbol{\lambda} = \sqrt{M} \text{DFT}\{[\mathbf{w}^T \ \mathbf{0}^T]^T\}$ . Note that the characterization developed in §3.1.2 concerning the null elements of  $\mathcal{Y}$  no longer applies to (5.4), due to the weighting of the exponentials by the components of  $\mathbf{h}$ . Nonetheless, if we suppose that  $\mathbf{h}$  is known *a priori*, we can still develop an AAS procedure to solve (5.4) for the eigenvalues  $\lambda_m$ , from which the impulse response  $\mathbf{w}$  can be recovered.

For simulation purposes, we consider the case where  $P = 3$  and  $L_w = L_h = 3$ . The elements of  $\mathbf{w}$  and  $\mathbf{h}$  are independently drawn from a standard Gaussian distribution, except for  $[\mathbf{w}]_1$ , which is set as  $[\mathbf{w}]_1 = 1$  to eliminate the scaling ambiguity. (Consequently, (5.4) is nonzero with probability one for all  $(m_1, m_2, m_3)$ .) Then, we construct  $\mathcal{K}$  as in (5.3), from which the data tensor is obtained via  $\mathcal{D} = \mathcal{K} + \sigma_N \mathcal{N}$ , where  $\sigma_N > 0$  controls the SNR and  $\mathcal{N}$  is a symmetric tensor whose elements  $n_{m_1, m_2, m_3}$  with indices satisfying  $m_1 \leq m_2 \leq m_3$  are independently drawn from a zero-mean Gaussian distribution, while the others are set by symmetry. To evaluate the outcome of a method, we measured the quantity  $\text{NMSE}_{\mathbf{w}} = \|\mathbf{w} - \hat{\mathbf{w}}\|_2^2 / \|\mathbf{w}\|_2^2$ , where  $\hat{\mathbf{w}}$  is the estimate of the filter  $\mathbf{w}$ , and also the reconstruction error  $\text{NMSE}_{\mathcal{K}}$ , which is calculated in the same manner as that described in §5.2.

We evaluate in this scenario all methods of Table 2, except for those involving the CALS algorithm, since (5.3) is not a CCPD. The employed AAS procedure consists in computing  $\lambda_1$  from  $y_{1,1,1}$  and then recovering each other eigenvalue  $\lambda_m$  using the equation for  $y_{m,1,1}$ . Regarding LMAS, a specialization of the algorithm to solve the equations (5.4) is used. The algorithms CPTOEP and TPARAFAC1 are applied to estimate a single circulant factor, which is only possible because  $\mathbf{W}$  is tall (and banded, in the case of TPARAFAC1). As for ALS,  $\hat{\mathbf{w}}$  is extracted from the column that, among all columns of all factors, provides the estimate which minimizes  $\text{NMSE}_{\mathbf{w}}$ .

We repeated the experiment for 1000 realizations, with the SNR varying in [20, 80] dB, and discarded the 1% worst realizations in terms of  $\text{NMSE}_{\mathbf{w}}$  for each method and each SNR level. The obtained results are displayed in Table 5, in which the lowest error(s) are again underscored. Observe that the results of ALS and LMAS with random initialization are again quite unsatisfactory, which highlights the difficulties faced in this nonlinear optimization problem. Despite them, the combination of these algorithms with the non-iterative ones produces very accurate estimates, whose errors are generally close together, except for some degradation of AAS-LMAS at 20 dB. As in §5.2, CPTOEP and TPARAFAC1 provide more accurate initial estimates than AAS, but at a higher computational cost.

TABLE 5  
Simulation results for the Wiener-Hammerstein identification scenario.

SNR (dB) Method	NMSE <sub>w</sub> (dB)					NMSE <sub>K</sub> (dB)				
	20	30	40	50	60	20	30	40	50	60
ALS	-23.7	-23.8	-23.8	-24.3	-23.7	-16.4	-16.7	-16.9	-17.4	-16.9
LMAS	-3.9	-4.4	-5.9	-2.8	-4.9	20.4	15.4	15.6	25.1	15.7
AAS	-4.1	-9.4	-16.7	-25.8	-35.7	9.7	1.0	-7.4	-18.1	-28.2
AAS-ALS	-31.5	-42.1	-52.2	-61.2	-68.6	-24.0	-34.5	-44.7	-53.6	-61.0
AAS-LMAS	-27.1	<u>-43.9</u>	<u>-53.9</u>	-63.9	-73.9	-22.3	-36.3	-46.3	-56.4	-66.4
CPTOEP	-24.9	-35.9	-46.1	-56.1	-66.2	-17.2	-28.3	-38.4	-48.5	-58.5
CPTOEP-ALS	-32.0	-42.5	-53.4	<u>-64.2</u>	<u>-74.2</u>	-24.5	-35.0	-45.9	<u>-56.9</u>	<u>-66.9</u>
CPTOEP-LMAS	<u>-33.9</u>	<u>-43.9</u>	<u>-53.9</u>	-63.9	-73.9	<u>-26.3</u>	<u>-36.4</u>	<u>-46.4</u>	-56.4	-66.4
TPARAFAC1	-14.6	-32.8	-43.1	-53.1	-63.1	9.0	-25.6	-35.6	-45.5	-55.5
TPARAFAC1-ALS	-31.9	-42.4	-53.2	-63.9	-73.9	-24.4	-34.9	-45.7	-56.5	-66.5
TPARAFAC1-LMAS	<u>-33.9</u>	<u>-43.9</u>	<u>-53.9</u>	-63.9	-73.9	<u>-26.3</u>	<u>-36.4</u>	<u>-46.4</u>	-56.4	-66.4

**6. Concluding remarks.** In this paper, we have proposed a novel approach for the problem of estimating circulant factors of a CCPD. This approach is based on the resolution of a system of homogeneous monomial equations which were derived by taking the multidimensional discrete Fourier transform of the tensor to be decomposed. We have shown that, under the standard Kruskal’s uniqueness condition, all solutions of these equations are related with each other in a simple way. We have also considered the symmetric CPD with circulant factors, whose resulting monomial equations are simpler and admit a finite number of solutions under the same uniqueness condition.

Numerical simulations were presented, where a simple *ad-hoc* solution of the monomial equations is compared with estimates given by several methods—this includes two existing non-iterative algorithms, a LM scheme for solving the monomial equations and a version of the ALS specialized for the CCPD, named CALS, which we propose in the present paper. Such *ad-hoc* solution is inexpensive to calculate, but is degraded in the presence of noise, since (i) it exploits only the minimum number of equations required and (ii) it resorts to divisions and multiplications to eliminate variables, which can produce large estimation errors. Nonetheless, by refining this solution with iterative methods, a much better performance is attained than when those methods are randomly initialized. This suggests that the AAS estimate often lies in the region of attraction of a global optimum (note that multiple global minima exist, due to the inherent indeterminacies of the CP decomposition), to where an iterative algorithm can then converge, thereby avoiding local minima. Indeed, this combination led to a noteworthy performance.

It should be emphasized, however, that our use of an *ad-hoc* procedure was mainly intended for validating the proposed approach. Conceivably, more sophisticated methods for the resolution of the monomial equations could lead to better results, which remains as a topic for future research. The authors envisage also an extension of the proposed approach for computing other similarly structured CPDs, which brings the possibility of addressing other practical problems. An example is the decomposition of the third-order cumulant tensor derived in [17, Sec. 4], which can provide an estimate of the impulse response of a linear communication channel.

## REFERENCES

- [1] E. ACAR, C. AYKUT-BINGOL, H. BINGOL, R. BRO, AND B. YENER, *Multiway analysis of epilepsy tensors*, Bioinformatics, 23 (2007), pp. i10–i18.
- [2] B. W. BADER, M. W. BERRY, AND M. BROWNE, *Discussion tracking in Enron email using PARAFAC*, in Survey of Text Mining II, Michael W. Berry and Malu Castellanos, eds., Springer London, 2008, pp. 147–163.
- [3] J. BRACHAT, P. COMON, B. MOURRAIN, AND E. TSIGARIDAS, *Symmetric tensor decomposition*, Linear Algebra and its Applications, 433 (2010), pp. 1851–1872.
- [4] R. BRO, *PARAFAC: Tutorial and applications*, Chemometrics and Intelligent Laboratory Systems, 38 (1997), pp. 149–171.
- [5] R. BRO AND H. A. L. KIERS, *A new efficient method for determining the number of components in parafac models*, Journal of Chemometrics, 17 (2003), pp. 274–286.
- [6] J. D. CARROLL AND J.-J. CHANG, *Analysis of individual differences in multidimensional scaling via an  $n$ -way generalization of “Eckart-Young” decomposition*, Psychometrika, 35 (1970), pp. 283–319.
- [7] J. D. CARROLL, S. PRUZANSKY, AND J. B. KRUSKAL, *Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters*, Psychometrika, 45 (1980), pp. 3–24.
- [8] P. COMON, G. GOLUB, L.-H. LIM, AND B. MOURRAIN, *Symmetric tensors and symmetric tensor rank*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1254–1279.
- [9] P. COMON, X. LUCIANI, AND A. L. F. DE ALMEIDA, *Tensor decompositions, alternating least squares and other tales*, Journal of Chemometrics, 23 (2009), pp. 393–405.

- [10] P. COMON, M. SORENSSEN, AND E. TSIGARIDAS, *Decomposing tensors with structured matrix factors reduces to rank-1 approximations*, in IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, Texas, USA, 2010, pp. 3858–3861.
- [11] J. P. C. L. DA COSTA, F. ROEMER, M. HAARDT, AND R. T. DE SOUSA JR, *Multi-dimensional model order selection*, EURASIP Journal on Advances in Signal Processing, 26 (2011), pp. 1–13.
- [12] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1084–1127.
- [13] D. E. DUDGEON AND R. M. MERSEREAU, *Multidimensional Digital Signal Processing*, Prentice Hall, NJ, USA, 1984.
- [14] G. FAVIER, T. BOUILLOC, AND A. L. F. DE ALMEIDA, *Blind constrained block-Tucker2 receiver for multiuser SIMO NL-CDMA communication systems*, Signal Processing, 92 (2012), pp. 1624–1636.
- [15] G. FAVIER, A. Y. KIBANGOU, AND T. BOUILLOC, *Nonlinear system modeling and identification using Volterra-PARAFAC models*, International Journal of Adaptive Control and Signal Processing, 26 (2012), pp. 30–53.
- [16] C. A. R. FERNANDES, G. FAVIER, AND J. C. M. MOTA, *Parafac-based channel estimation and data recovery in nonlinear MIMO spread spectrum communication systems*, Signal Processing, 91 (2011), pp. 311–322.
- [17] C. E. R. FERNANDES, G. FAVIER, AND J. C. M. MOTA, *Blind channel identification algorithms based on the Parafac decomposition of cumulant tensors: the single and multiuser cases*, Signal Processing, 88 (2008), pp. 1382–1401.
- [18] R. M. GRAY, *Toeplitz and circulant matrices: A review*, Foundations and Trends in Communications and Information Theory, 2 (2006), pp. 155–239.
- [19] R. A. HARSHMAN, *Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-modal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
- [20] C. J. HILLAR AND L.-H. LIM, *Most tensor problems are NP-hard*, Journal of the ACM (JACM), 60 (2013), pp. 45:1–45:39.
- [21] A. Y. KIBANGOU AND G. FAVIER, *Non-iterative solution for PARAFAC with a Toeplitz matrix factor*, in EUSIPCO, Glasgow, UK, 2009, pp. 691–695.
- [22] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM review, 51 (2009), pp. 455–500.
- [23] J. B. KRUSKAL, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and its Applications, 18 (1977), pp. 95–138.
- [24] J. LEBRUN, *Normal forms in statistical signal processing*, in Gröbner Bases in Control Theory and Signal Processing, de Gruyter, 2007, pp. 107–126.
- [25] M. MARCUS, *Finite dimensional multilinear algebra, Part 1*, Pure and applied mathematics, M. Dekker, New York, NY, 1973.
- [26] D. W. MARQUARDT, *An algorithm for least squares estimation of non-linear parameters*, SIAM Journal, 11 (1963), pp. 431–441.
- [27] M. MØRUP, *Applications of tensor (multiway array) factorizations and decompositions in data mining*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1 (2011), pp. 24–40.
- [28] P. PAATERO, *Construction and analysis of degenerate PARAFAC models*, Journal of Chemometrics, 14 (2000), pp. 285–299.
- [29] N. D. SIDIROPOULOS AND R. BRO, *On the uniqueness of multilinear decomposition of N-way arrays*, Journal of Chemometrics, 14 (2000), pp. 229–239.
- [30] N. D. SIDIROPOULOS, G. B. GIANNAKIS, AND R. BRO, *Blind PARAFAC receivers for DS-CDMA systems*, IEEE Transactions on Signal Processing, 48 (2000), pp. 810–823.
- [31] M. SORENSSEN AND P. COMON, *Tensor decompositions with banded matrix factors*, Linear Algebra and its Applications, 438 (2013), pp. 919–941.
- [32] M. SORENSSEN AND L. DE LATHAUWER, *Blind signal separation via tensor decomposition with Vandermonde factor: canonical polyadic decomposition*, IEEE Transactions on Signal Processing, 61 (2013), pp. 5507–5519.
- [33] A. STEGEMAN, *Low-rank approximation of generic  $p \times q \times 2$  arrays and diverging components in the Candecomp/Parafac model*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 988–1007.
- [34] G. TOMASI AND R. BRO, *A comparison of algorithms for fitting the PARAFAC model*, Computational Statistics and Data Analysis, 50 (2006), pp. 1700–1734.